Response to reviewer 1 Comments

ERA5 reanalysis data is used to evaluate the performance of two seasonal forecast model systems, the SEAS5 and MF7, in simulating the variability of the Saharan Heat low (SHL) at various timescales over the 1993 – 2016 period. Strengths and weaknesses of each forecast model are discussed, and statistical bias correction methods are applied to improve the forecast of the SHL in the forecast models. While the methods applied in this paper seem appropriate for the analysis conducted, I do have a serious concern (see below) that needs to be addressed.

We first thank the reviewer for his availability and interest to evaluate the work.

My concern is that the entire validity of the results presented rely on the assumption that the ERA5 reanalysis is providing reliably accurate information since it is being used here as the "target" for comparison and the bias correction being applied. In my opinion this may not be a sound assumption to make over a relatively remote region of the Sahara where there are far fewer observations constraining the ERA5 reanalysis. This means that there certainly is some degree of uncertainty in the ERA5 reanalysis data, but the authors do not explicitly address this uncertainty anywhere in the manuscript. What is needed is for the analysis to be expanded such that not just ERA5 is used as a "target", but also other atmospheric reanalyses (e.g., such as JRA-55 and/or MERRA2, to name a couple) are also evaluated. In doing so, results from multiple reanalyses can be compared and explicitly discussed to address this uncertainty and provide a means to talk about the greater robustness of the findings in general.

We agree with the reviewer that ERA5 may present uncertainties compared to surface-based observations in the area of interest. Nevertheless, the assimilation of satellite observations as well as operational ground-based observation and radiosounding data available on the fringes of the Saharan Heat Low (southern and western Algeria, southern Mali, Mauritania, Morocco) in the ECMWF IFS contributes to improving the quality over the Saharan region where the observations are few. In consequence, ERA5 has shown good skill to represent large scale atmospheric patterns and circulations in the area (Olauson 2018; Ramon et al 2019); it is commonly used in West Africa [Diouf et al 2019; Guigma et al 2020, 2021; Osei et al 2021].

References

Olauson, J. (2018). ERA5: The new champion of wind power modelling?. *Renewable energy*, *126*, 322-331.

Ramon, J., Lledo, L., Torralba, V., Soret, A., & Doblas-Reyes, F. J. (2019). What global reanalysis best represents near-surface winds?. *Quarterly Journal of the Royal Meteorological Society*, *145*(724), 3236-3251.

Diouf, D., Niang, A., & Thiria, S. (2019). Deep Learning based Multiple Regression to Predict Total Column Water Vapor (TCWV) from Physical Parameters in West Africa by using Keras Library. *arXiv preprint arXiv:1912.07388*.

Guigma, K.H., Todd, M. & Wang, Y. Characteristics and thermodynamics of Sahelian heatwaves analysed using various thermal indices. *Clim Dyn* **55,** 3151–3175 (2020). https://doi.org/10.1007/s00382-020-05438-5

Guigma, K.H., Guichard, F., Todd, M. *et al.* Atmospheric tropical modes are important drivers of Sahelian springtime heatwaves. *Clim Dyn* **56,** 1967–1987 (2021). https://doi.org/10.1007/s00382-020-05569-9

Osei, M. A., Amekudzi, L. K., & Quansah, E. (2021). Characterisation of wet and dry spells and associated atmospheric dynamics at the Pra river catchment of Ghana, West Africa. *Journal of Hydrology: Regional Studies*, *34*, 100801.

Without the inclusion of this expanded analysis the findings only have limited value because they are not placed in a broader context. Unfortunately, what I am suggesting above will likely result in a large reworking of the entire manuscript and will take some time to complete. Thus, for this reason I have recommended to reject and resubmit for this manuscript. I encourage the authors to do so because there is good potential to advance our weather prediction capabilities from a study such as this one. Below are some additional comments I had for the authors as they update their manuscript accordingly.

We are sorry to disagree with the reviewer statement. First of all, there is a large body of literature on forecasting that uses only one type of reanalysis (or satellite observation) to assess the quality of forecasts [Landman et al 2012, Pepler et al 2015, Batté et al 2018, Carrão et al 2018, Lavaysse et al 2019, Pirret et al 2020, Vogel et al. 2021]. Moreover, given the existing large body of literature showing the robustness of ERA5 reanalyses in Africa -for model validation and process studies- it seems quite relevant to use them. Adding new reanalysis datasets without detailed study and using complex and rich sets of ground observations does not necessarily mean an improvement of the knowledge of the observed situation. The proposed

additional work, although very interesting, is beyond the scope of the present study. Furthermore, the use of multiple sets of reanalyses would make the interpretation of the quality of the forecasts very difficult if not impossible.

Nevertheless, some additional work, relevant to the reviewer's comment was conducted to compare the characteristics of the SHL as seen by ERA5 with those derived from the MERRA2 reanalyses (which has also shown its ability to represent the major components of the African monsoon). This work has been carried out and has been integrated into the supplementary material of the study. Thus the forecast scores of MF7 and SEAS5 presented in that study can be now compared to the differences between the two sets of robust reanalyses and so what we could consider a proxy of their uncertainties.

References

Landman, W. A., & Beraki, A. (2012). Multi-model forecast skill for mid-summer rainfall over southern Africa. *International Journal of Climatology*, *32*(2), 303-314.

Pepler, A. S., Díaz, L. B., Prodhomme, C., Doblas-Reyes, F. J., & Kumar, A. (2015). The ability of a multi-model seasonal forecasting ensemble to forecast the frequency of warm, cold and wet extremes. *Weather and Climate Extremes*, 9, 68-77.

Batté, L., Ardilouze, C., & Déqué, M. (2018). Forecasting West African heat waves at subseasonal and seasonal time scales. *Monthly Weather Review*, *146*(3), 889-907.

Carrão, H., Naumann, G., Dutra, E., Lavaysse, C., & Barbosa, P. (2018). Seasonal drought forecasting for Latin America using the ECMWF S4 forecast system. *Climate*, *6*(2), 48.

Lavaysse, C., Naumann, G., Alfieri, L., Salamon, P., & Vogt, J. (2019). Predictability of the European heat and cold waves. *Climate Dynamics*, *52*(3), 2481-2495.

Pirret, J. S., Daron, J. D., Bett, P. E., Fournier, N., & Foamouhoue, A. K. (2020). Assessing the skill and reliability of seasonal climate forecasts in Sahelian West Africa. *Weather and Forecasting*, *35*(3), 1035-1050.

Vogel, P., Knippertz, P., Gneiting, T., Fink, A. H., Klar, M., & Schlueter, A. (2021). Statistical forecasts for the occurrence of precipitation outperform global models over northern tropical Africa. *Geophysical Research Letters*, *48*(3).

- Lines 21 – 23: sentence is worded awkwardly and can be misinterpreted. How it is currently written implies the only reason the SHL emerged as a key component of the WAM system is because of AMMA, which of course is not true. Suggest the authors update the text to better clarify that the AMMA project significantly highlighted the importance of the SHL in influencing variability of the WAM system

We reformulated the sentence according to the reviewer's suggestion so that it appears more clear. {" During the AMMA project (Redelsperger et al., 2006), the SHL has been used as a key component to assess the variability of the WAM system."}

- Line 34: Typo: Thorncroft and M. 1999 Need to fix this as you appear to be missing the second author's last name both here and in the references.

We have fixed this issue according to the reviewer's comment.

- Line 76 – 77: "very hot temperature" Can you provide a range of temperatures here to show what you mean comparable to what you did for RH later in the same sentence?

We have clarified that point by indicating in the Sahara region, mean temperature values are sometimes over 30°C and mean maximum temperature values during summer (May to September) are over 40°C.

- Line 87 – 91: "… detected the SHL with occurrence of more than 70% during the boreal summer, …." Using what data? Daily? Hourly? Can you provide more information here about what you mean from all these prior studies that you presumably are taking the same regions? Likewise, you discuss detecting the SHL here, but you have not yet mentioned exactly how you plan to detect the SHL. What metric(s) are you using? I presume this is coming a little later, but maybe it should come first.

We have clarified this point by adding these sentences.

{"The SHL has been detected by Lavaysse et al 2009 over the Central SHL with more than 70% of occurence in boreal summer using ERA-40 daily reanalysis. This detection was done by using the low level atmospheric thickness (LLAT, i.e. the atmospheric thickness between two geopotential levels at 925 to 700 hPa). In Lavaysse et al. (2016), it has been shown that the temperature field at 850 hPa can be used as a proxy of the LLAT. The choice of the 4 regions was supported by previous studies: Lavaysse et al 2009 highlight a maximum activity of the SHL in the

CSHL location; Roehrig et al 2011 show that the SHL tends to migrate from the West to the East during the season which explain the WSHL and ESHL locations. The detection of the SHL is presented in section 2.4.1, but according to the reviewer comment we will re-organise the section and put it first before showing the SHL boxes."}

- Line 95: Is the daily temperature just for a specific level/levels? If so, which? Again – this is related to my other comment earlier that it may be better to explain how you intend to detect the SHL before the discussion in 2.1 and 2.2.

We thank the reviewer for this relevant comment. We have clarified this point by explaining we use the daily temperature at 850 hpa. We added this information in the text to be more explicit and re-organised the sections so that the readers get the information on detection first. {" The ERA5 atmospheric variable studied here is daily temperature at 850 hpa with a spatial resolution of 0.25° x 0.25° downloaded on the climate data store website: https://cds.climate.copernicus.eu/. "}

- Lines 132 – 135: What data was the Lavaysse (2015) using (certainly not ERA5), and have you confirmed that it is indeed valid for ERA5 and the MF7 and SEAS5? It would be helpful to convey this explicitly to build confidence in your methodology here.

Lavaysse et al. (2016) used ERA-Interim reanalysis and showed a correlation between the temperature at 850 hpa and the LLAT for the detection of the SHL. As ERA5 is an improvement of ERA-Interim, we assume that the correlation between T850 hpa and the LLAT will be preserved in ERA5 (see previous comment). We suppose this is also true for the forecast models. We added this information in the text and reformulated lines 132-135 as follow:

{"Lavaysse et al. (2016) using ERA-Interim reanalysis, showed that the 850 hPa temperature field is well correlated to the LLAT and can be used as a proxy for the monitoring of the SHL (detection and intensity). As ERA5 is an improvement of ERA-Interim, we assume that the correlation between T850 hpa and the LLAT is preserved in ERA5. We suppose this is also true for the forecast models. Consequently in this study, we use the temperature at 850 hPa to analyse the SHL characteristics. Because fixed boxes are used, the detection of the SHL is not needed, but, strong (weak) phases of the SHL will be associated with high (low) respectively temperatures."}

- Lines 222 – 226: In Figure 1 and other figures with shading (Figs. 4, 6, 7) there is not enough contrast between the different color hues making it hard to visually interpret values from the figure. Thus, it is hard to evaluate how well

SEAS5 and MF7 are doing compared to ERA5. Recommend the authors improve the figures by increasing the contrast between the color values used and possibly add line contours to label interval levels.

We improved the contrast between the color values in the figures to make the understanding of our results easier for the community. We thank the reviewer for the suggestion.

- Line 223 – 225: I don't understand what the authors mean by "coherent climatologies of the SHL over the Sahara". I think they mean that the SEAS5 and MF7 reasonably replicate the climatology of ERA5, but I am not certain. Please clarify.

Yes, the reviewer is right by "coherent climatologies of the SHL over the Sahara ", we mean that SEAS5 and MF7 are able to reproduce the climatology of the SHL over the Sahara. We reformulated the sentence to be more clear.

- Line 225 – 226: It is unclear what is meant by "A progressive decrease in the intensity of the SHL is also observed over the North of Libya". MF7 does not appear to yield the relatively cooler temperatures over northeastern Africa that are shown in ERA5. Is this what is meant? In any case the authors need to clarify this comment better.

Not exactly, by "A progressive decrease in the intensity of the SHL is also observed over the North of Libya", we want to highlight the fact that in all the 3 products, we observe a diminution in the intensity of the SHL over the North of Libya [Fig. 1]. This feature is very marked in ERA5 and SEAS5, and a little bit in MF7. We reformulated the sentence to be more explicit according to the reviewer's comment: {" In all the 3 products, a progressive decrease in the intensity of the SHL is observed over the North of Libya [Fig. 1]; this feature is very marked in ERA5 and SEAS5, and a little bit in MF7."}

- Line 237 – 239: " … to get a robust selection of events at different periods." Can you explain more explicitly what is meant by robust selection here? Also – it would be helpful if the authors would explicitly mention with a sentence or two in the manuscript how the distributions change when the arbitrary threshold changes from 0.5 to 10.

The distribution of events at different periods has been assessed through the sensitivity test on several thresholds from 0.5 to 10. The analysis of the results in terms of significant days (days associated with an intensity of signal greater than a given threshold) reveals that the signal is more intense for a threshold value of 1 compared

to other threshold values. So, we mean by "robust selection of events" here the process which consists in selecting predominant events. We clarified that point in the new version of the study and reformulated as follow:

{" This threshold of 1 has been selected arbitrarily after applying a sensitivity test on several threshold values from 0.5 to 10 to focus on predominant events at different periods. We noticed globally a decrease of events occurrence with high threshold values of the spectral power. Note that the sensitivity to the threshold values does not significantly impact our results (not shown)."}

- Line 240: " …. in all our products …." By products, do you mean seasonal model forecasts? Suggest clarifying to appeal more to readers less familiar with the seasonal forecasting lingo.

By " …. in all our products ….", we mean here in all the datasets used for the study: the reanalysis ERA5 and the seasonal forecast models (SEAS5 and MF7). We clarified that point in the manuscript to avoid confusion: {" We observe a similar behaviour in ERA5, SEAS5 and MF7 in terms of significant days with an increasing number of days with periods up to 10days followed by a quite steady activity for longer periods."}

- Lines 277 – 279: This seems speculative. Given that you have all the output you could nail down whether or not this is what is happening.

We agree with the reviewer that the origin suggested of the hot bias present in MF7 over the eastern part of Sahara is speculative. We could make some investigations to highlight the real cause of this behaviour in MF7, but this set of analysis requires more knowledge about the physical processes occurring in that area, which is out of the scope of this study.